DOCUMENT RESUME

ED 239 005 UD 023 308

AUTHOR Miller, Norman; Carlson, Michael
TITLE School Desegregation as a Social Reform: A Meta-Analysis of Its Effects on Black Academic Achievement.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Dec 82
NOTE 76p.; Paper submitted as one of a collection from the National Institute of Education Panel on the Effects of School Desegregation. For related documents, see UD 023 302-307.
PUB TYPE Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Achievement Gains; *Black Students; *Desegregation Effects; Effect Size; Elementary Secondary Education; Mathematics Achievement; Meta Analysis; Outcomes of Education; *Performance Factors; Program Effectiveness; Program Evaluation; Racial Integration; *Racial Relations; Reading Achievement; Research Reports; School Desegregation; *Self Esteem; Social Change; White Students

ABSTRACT

Although school desegregation was initiated to address a social inequity--segregated schooling was seen as stigmatizing blacks as a social group--research has focused primarily on desegregation's effects on black academic achievement and self-esteem. Two problems have made this research difficult: the ambiguity of the term "school desegregation" and the quality and characteristics of the research designs used to study it. In this meta-analysis of 19 desegregation studies prepared for the National Institute of Education, the effect size method is used. Results show that the effects of desegregation on verbal tests is significant as is the pooled verbal and math effect size, but the math test effect size is not significant. Analysis of white achievement gains in three of the studies shows that black gains relative to white gains are small, thus suggesting that black gains are not attributable to desegregation per se. Other factors affecting academic outcomes in desegregated settings--anxiety and threat, self-concepts and aspirations, peer comparison, expectations, peer relations, school effects, teachers, and students--have diverse effects on and are affected in diverse ways by desegregation. Although desegregated schooling has only a moderate positive effect on black achievement, desegregation is nevertheless a requisite if the social issue of interracial acceptance is to be addressed. (CMG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED239005

School Desegregation as a Social Reform: A Meta-Analysis of its Effects on Black Academic Achievement*

Norman Miller
University of Southern California

Prepared for the National Institute of Education
December, 1982

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

* Written with the assistance of Michael Carlson. I also wish to acknowledge the assistance of Rolf Holtz in preparing the references and the Social Science Research Institute, University of Southern California, in preparing the text.

UD023308

Paper submitted as one of a collection from the National Institute of Education Panel on the Effects of School Desegregation

School Desegregation as a Social Reform:

A Meta-Analysis of its Effects on Black Academic Achievement

Norman Miller

University of Southern California

This paper addresses the specific question of what effect school desegregation has had on the achievemment test scores of black children. It is one of a common set of papers addressing this issue, all prepared for the National Institute of Education. All of the papers base their conclusion and analysis on the same set of core studies that the panel of experts, selected by NIE to perform the review task, have agreed upon as meeting certain criteria for inclusion among those to be reviewed.

Before summarizing the results of these core studies, it is important first to put the question itself into an historical context, and second, to discuss the criteria for inclusion and exclusion of studies and the procedures used in performing the analysis. Then, after presenting their findings, their meaning and policy implications will be discussed.

Background

School desegregation was initiated to address a social inequity--the impairment of minority children's right to equal educational opportunity. The Brown decision required school desegregation as a remedy for prior discrimination, declaring separate facilities inherently unequal. It is important to note that in the view of Brown, educational outcome is not the issue. Had it been shown that blacks in segregated schools performed on

standardized achievement tests as well as did whites in segregated schools, inequality of educational opportunity would nevertheless prevail according to Brown. This is not to deny that the evidence of social scientists that was presented in the case did focus on inequalities between black and white children in their self-concepts, motivation, and academic performance. In its ruling, however, the court seemed concerned primarily with the notion that segregated schooling ineluctably stigmatized blacks as a social group.

> "Does segregation of children in public schools solely on the basis of race, even though the physical facilities and other 'tangible' factors may be equal, deprive the children of the minority group of equal educational opportunities? We believe that it does . . . to separate Negro school children from others of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hearts and minds in a way unlikely ever to be undone . . . in the field of public education the doctrine 'separate but equal' has no place. Separate educational facilities are inherently unequal. ...
>
> Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law; for the policy of separating the races is usually interpreted as denoting the inferiority

of the Negro group" (Brown v. The Board of Education, 1954).

The fact of educational separation was the problem to be cured; the cure was desegregation. In principle, this logic is simple and straightforward; it requires no other major ingredients (such as, for instance, proof that desegregation will eliminate or reduce wage inequities, or other specific differences in the outcomes of blacks and whites). Of course, when school desegregation was implemented in specific cities and school districts, the method and degree of desegregation became important issues. Presumably, in court mandated plans, the extensiveness of a court imposed remedy should in some degree correspond to the severity or magnitude of the acts that created segregated schooling (Black, 1960; Kluger, 1977).

Americans are basically sympathetic to the plight of blacks. They know that despite the beneficial social changes for blacks that have occurred over past decades, discrimination exists and most believe it wrong. Most believe that the full weight of the Federal government should be martialed in order to eliminate such injustice. Two decades ago 91 percent of whites favored equal voting rights, 87 percent favored the right to a fair jury trial and to nonsegregated public transportation, and 72 percent favored integrated education. Despite the fact that white Americans by a margin of 2 to 1 felt in 1966 that black children would not be better educated in integrated classrooms, they had no deep aversion to black children attending the same school as their own offspring. By a margin greater than 3 to 1, they denied that the

education of white children would suffer if blacks are in their classroom. Three out of four white Americans approved of the Court ruling outlawing segregation in education (Brink & Harris, 1966, p. 131). There is, of course, substantial slippage between belief and action. Despite this endorsement of the moral aspects of court rulings, most whites may not be inclined to do anything specific about helping to bring about integration in schools.

In viewing the courts' position, legal scholars have noted that the remedy or restitution (viz. desegregation) was often imposed on parties other than either the perpetrators of segregation (for instance, the school board that created it) or on their victims (those who graduated from the segregated school system). This characteristic of legally imposed remedies has led some legal analysts to interpret the underlying legal principle or goal not as restitution to the injured party, but instead, as group protection. Child labor laws or minimum age drinking laws might be other instances of the same principal. For a discussion of this view, see Yudof's (1980) interpretation and discussion of Dworkin (1970).

Since the time of Brown, social science seems to have concerned itself with the specific effects of desegregated schooling on black academic achievement, black self-concepts, and on interracial hostility and prejudice. Although these three issues were prominent in the social science statement appended to Brown, they are not the same as racial separation and stigmatization. Among the three, the one that most closely approaches stigmatization in meaning, or is most directly related

to it, is intergroup hostility and prejudice. It should be noted, however, that hostility and prejudice do not necessarily denote stigmatization. Although ingroup bias is ubiquitous in intergroup relations, not all or even most outgroups are stigmatized. We frequently encounter outgroups in our daily lives. Common examples of reciprocal ingroup-outgroup pairs might be: production and sales personnel in a particular manufacturing company; two fraternities on a university campus; two teams in a baseball little league; members of opposing political parties; etc. Yet ordinarily, none of these groups are stigmatized by each other.

The point here is that the issues that have concerned social scientists, namely, low academic achievement and poor self-concepts among black children, if not prejudice as well, are not the causes of stigmatization. As implied by Campbell's argument, even if the directions of existing difference were reversed, stigmatization would persist (Campbell, 1967). The flexibility of our evaluative terminology allows <u>any</u> direction of difference to be positively labeled when describing ingroup members and negatively labeled when depicting outgroups. ("We are firm; they are pigheaded"). Thus, to the extent that racial-ethnic differences in academic achievement and self concept exist, it makes more sense to view them as consequences than as causes of stigmatization. And if they are consequences, they certainly are not the only ones. Other possible consequences are wage inequities, inequalities in employment rates, lower voter turnout

among blacks, higher death and disease rates, etc.

Social Science Research on School Desegregation

In their research on school desegregation why have social scientists focused their attention primarily on its effects on black academic achievement and black self-esteem? Perhaps in part they took their instruction from the emphasis found in the social science statement that was appended to the plaintiffs' case in Brown, which put impairment of black childrens' self-concept as the most pivotal or central consequence of black stigmatization, and viewed other consequences as flowing from or being caused by this key deficiency (Stephan, 1978).

The fact that studies of the effect of school desegregation on academic achievement, however, are so much more prevalent than those of any other variable reflects two additional factors. First, it undoubtedly reflects the fact that measures of academic achievement are so routinely administered by school districts. Second, such measures are very readily seen as central to the educational mission. This makes such studies more appealing to administrators who must approve the researcher's intrusion into school activities and/or records, but also, to the public as well.

The courts too, seem to have been responsive to this manifest connection. Despite the fact that some research suggests that education contributes relatively little to one's life outcomes (Jencks, Smith, Bane, Cohen, Gintis, Heynes, & Michelson, 1972) the California State Supreme Court (Crawford, 1975) viewed desegrated education as a means of increasing the social mobility

of minorities, presumably by providing better education and higher levels of cognitive mastery to minority students. Yet, Cook (1979), who was one of the authors of the social science statement appended to Brown, states that it "nowhere predicted improvement in the school achievement of black children as a consequence of desegregation" (Cook, 1979). Nevertheless, it is clear that courts as well as social scientists, have been interested not merely in the fact of segregated schooling, but also, in the effects of desegregated schooling on minority children.

Two problems have made it difficult for social scientists to provide answers about the effect of school desegregation. The first is the ambiguity in the meaning of the term "school desegregation." The second stems from the quality and charactistics of the research designs used to study it.

The definition of school desegregation. At first thought, the meaning of the term "school desegregation" seems straightforward. An analysis of how school desegregation has been immplemented in any set of communities or cities, however, reveals substantial variability. Thus, the meaning of the term is in fact vague. The only common definitional element among studies of its effects is that the ratio of minority and white students in a classroom or school has been altered. By how much? Are the whites in a classroom more or less numerous than the blacks? Is the percentage of minority students in the class or school changed from 98 percent to 45 percent, 98 percent to 5 percent, or 55 percent to 45 percent? Are the changes in percentages made in all classes, or just at certain grade levels or programs within the

school? Are both groups of children shifted to new schools or is just one of the groups? Is the teacher familiar to one or both groups of students or do the students have a new and unfamiliar teacher? Do both groups retain friends from the previous year in their class? To what extent have other important factors other than the ratio of white to minority students also been altered (e.g., the curriculum, the student teacher ratio, the quality of physical facilities, the quality of teaching materials, the quality of teachers, etc.)?

The problems created by an ambiguous definition can be illustrated by an analogy. "Consider the question "Is <u>eating food</u> good for humans?" Although on first thought the answer is obviously "yes," we can quickly see that the answer will depend on what is eaten and how. If the chicken salad has "turned", or the plate it is served on is lead-contaminated then the answer becomes no. If a child is fed only an ounce of food three times a day or the food is merely rubbed on the child's stomach, it will starve. It might also starve if the only food available were unpalatable (e.g., half digested dog food taken from a dog's stomach). A nutritionally balanced high-protein drink may sustain life but also cause one's teeth to drop out. Extended hospitalization for malnutrition might give one bed sores.

The examples above are not the "ordinary" instances of eating. But what are the "ordinary" instances of school desegregation? There are numerous circumstances in which few would expect desegregated schooling to produce academic gains for blacks: e.g., when teachers, students, or principals in receiving

schools are prejudiced against blacks (the food is poisoned); when there is only one or two of them in a classroom, or when they are ignored in the classroom (too little food to provide nourishment); when the curriculum is not modified to match their current performance level, and consequently is not assimilated (food is rubbed on their stomach); when they are made to feel rejected and incompetent (the food is unpalatable). On the other hand, it may produce academic gains but, simultaneously, as a consequence of exposure to higher performing classmates, lower their academic self-concepts (bed sores).

Americans may feel it is better or more moral to ship government overstocks of potatoes to an undernourished third-world country than to dump them in the ocean. As we have learned in the past, however, shipping food to people is not the same as nourishing them. Potatoes won't help if they arrive rotten, or if the receiving country lacks adequate mechanisms for distributing them. Nor will they help if protein deficiency is the problem. But nevertheless, despite our failure to achieve the goal of nourishing a famine-plagued third world country we might feel righteous about our efforts.

Simply put, many factors are relevant to school outcomes. Those factors that go hand in hand with desegregation in one setting may not in the next. Consequently, the meaning of the term varies from one study to the next, and often, in ways that are important but not well documented.

Research designs in studies of school desegregation. As indicated, a second problem in assessing the effects of school

desegregation is that researchers have rarely used a methodology that permits inferences about what it was that caused some observable difference between the comparison groups (segregated and desegregated students). This issue is quite separate from the previous one, which pointed to the variation in the meaning of the term desegregation and covariation of other factors with implementation of a change in the ratio of blacks to whites in a school. It refers instead to the fact that children, classrooms, or schools are almost never randomly assigned to comparison conditions. As a result, one cannot know whether initial differences between the groups account for (or cause) the differences found after the treatment (desegregated schooling).

Experts are agreed that attempts to select out from, (a) those students who continue to have segregated schooling and (b) those students who change to desegregated schooling, two subsets of children that are matched (or on the average equal) on key variables (e.g., IQ) will not solve the problems. If the so-called matched groups were measured again on the variables on which they were originally matched, they will again differ from each other in the direction in which they initially differed.* Similarly, they will also differ on variables correlated with the variable on which they were matched. Consequently, if, for instance, a high IQ implies better ability to learn and if, prior to their desegregation, the average IQ of the desegregated

---

*Technically termed regression, this effect is due to the fact that the measuring instruments (tests) do not tell us each person's true score; there is a component of error in each score.

students exceeded that of those who remained segregated, they might well perform better after desegregation. Such a difference might just as readily be attributed to the initial difference in IQ as to the difference in type of schooling. Why might students with higher IQ's naturally appear more frequently in the desegregated group? Parents and children who are brighter may be more motivated to seek out better schools. If they believe desegregated education to be superior, they will push to be in that program, to be included sooner in the desegregated group, or to be assigned to the desegregated school, etc., (e.g., Gerard & Miller).

Methodological Considerations for Summarizing the NIE Set of Studies

Procedures for Combining the Results of Studies

Several different methods exist for summarizing the outcomes of a group of studies. Recently these procedures have come to be called meta-analysis (Glass, 1976). One procedure is simply to tally the number of studies giving positive versus negative effects. This box score or voting approach is crude because it fails, for instance, to acknowledge differences among studies in the strength or magnitude of difference between comparison conditions. Almost no experts now advocate the voting method alone (Hunter, Schmidt, & Jackson, 1982). Furthermore, the voting or box score method can lead to erroneous conclusions due to "'false' conflicting results" in the literature (Hunter et al. p. 132).

The z-score method provides an alternative procedure for representing the size of the relationship betwen the treatment variable and the dependent measures in a given study. It requires computing the exact p of the statistic employed by the original researcher (and dividing it in half if a two-tailed test was employed) and then converting each p value to an exact z -score, based on the normal probability distribution. The sum of these z -scores across studies is then divided by the square root of the number of findings included to generate an overall z -score and its associated probability level. This provides an estimate of overall statistical significance, assessing the likelihood that the results of the entire pool of studies reflect chance outcomes. (This particular procedure typically understates significant effects because many authors do not include specific t, F, or χ^2 values in their research reports, and as a result nominal, rather than exact, p values have to be entered into the analysis.) With this method, a fail-safe n can be calculated to determine the number of additional studies with summed z -scores that total to zero that would be needed before the probability value associated with the overall z would exceed the .05 level.

The effect size method, is the most preferred method and the one used for this paper. In this method the difference between the means or pairs of treatment conditions in each study is divided by the within-group standard deviation of the outcome measure employed, thus yielding a standardized mean difference score (Glass, 1977). These difference scores can then be averaged across studies in order to generate an overall effect size

estimate.

Evaluating the Strength of Research Designs

Apart from generating summary estimates of overall effects, meta-analysis procedures can in principle be utilized to assess whether characteristics of research design and/or program implementation features are related to program effectiveness. For this purpose, characteristics of subjects, studies, and programs must be coded and then entered as predictors in multiple regression analyses, with estimates of size of effects as the dependent variable. Examples of such predictor variables might be factors such as age of program recipients, nature of the experimental design employed in the study, the extent of parental involvement in the program, etc. In general, the search for such predictor or moderator variables is highly prone to capitalization on chance unless the number of studies is very large. In the present case many statistical experts might judge the number of studies as too few to justify application of this procedure.

In the present case the study selection criteria imposed by the panel attempted to eliminate particularly weak studies from consideration. This does not mean that all or even most studies that survived the weeding out imposed by application of the minimum procedures are strong studies. They are not. And typically, studies with weak research designs show stronger or more positive effects than do those with stronger designs. For instance, in a meta-analysis of the larger body of school desegregation research concerned with achievement test performance, Krol (1978) found an average effect size of +0.21

among studies with weak designs, whereas among those with stronger designs, the effect was reduced by half (+0.10). While the effects of several design factors (threats to validity) have been found to be negligible in some educational contexts (Walberg, 1981), their influence nevertheless should be assessed whenever meta-analyses are undertaken in any new research arena. By imposing the selection criteria that we did, however, most of the variation in strength of design found in the total set of nineteen studies on school desegregation and academic achievement has been eliminated.

As indicated above, in addition to analyses involving research design considerations, it is ordinarily important to separate studies in terms of variables associated with the strength of program implementation. For this purpose, studies ideally should be rated or classified on implementation variables independently of knowledge of their outcomes. Unfortunately, the studies analyzed for this paper do not provide much information on correlates of (or strength of) the implementation or desegregation. Moreover, it is not even clear what "strength of implementation" means with respect to school desegregation.

Variation in Number and Type of Dependent Measure

In the subset of studies analyzed for this report the specific dependent measure varies from one study to the next. Not only do studies use different measures of verbal achievement, but within the same study the measure used prior to the implementation of desegregation may differ from that used later. In addition, some studies also include measures of achievement in mathematics,

science, and other subjects, as well as verbal achievement.

Does it make sense to try to summarize studies whose measures of verbal achievement differ from one study to to the next? It depends on the situation or problem. Although, for instance, it may make perfect sense to distinguish between vocabulary mastery and reading comprehension for some studies of educational success, in the present case there is little or no theoretical reason to expect school desegregation to differ in its impact on the two. In other words, with respect to the issue of whether school desegregation affects black academic achievement, different measures of verbal performance are conceptually interchangeable, in that they all tap some aspect of the verbal component of the academic curriculum.

For the same reason, the distinction between measures of verbal achievement and mathematical (and/or other academic areas such as science) can also be ignored, being merely another instance of the same issue; again, there appears to be little theoretical reason to think desegregation might affect the several areas of mastery differently. This line of reasoning argues that a single effect size be computed across studies regardless of variation across studies in the particular dependent measure (e.g., vocabulary, reading comprehension, mathematics, social studies, etc.).

In addition to variation among studies in their dependent measure, many studies report outcomes for several dependent measures. In this case, we are not dealing just with variation across studies in their dependent measure, but with multiple

outcomes on the same set of children. Here, the ideal procedure would convert the two sets of scores on each child (math and verbal achievement test score) to standard scores which would then be averaged for each child. The effect size for each study would then be computed on these averages. This results in each study contributing one value to the meta-analysis and at the same time minimizes error of measurement. Unfortunately, in the present instance this cannot readily be done because the raw score information is not available. To ignore the issue and treat the separate outcomes in math and verbal performance obtained in a single study as separate entries in the meta-analysis ignores the fact that these outcomes are not independent. Although not perfectly ideal, the best solution is to average the two effect sizes. This assures that studies with more measures are not given greater weight than those with few (or one).

Multiple Subject Groups

The same logic applies to the analysis of subgroups or multiple groups within the same study. The ideal procedure is to use an overall test across all subgroups. If this is not provided by the individual researcher, then the best alternative is to average the effect sizes computed for each subgroup.

Criteria for Inclusion

Appendix A lists the criteria agreed upon by the NIE panel as a basis for inclusion of studies to be analyzed. These yielded a core sample of 19 studies. Only studies included in the NIE core sample were considered appropriate for meta-analysis. This

requirement provides the first entry in Table 1, which details additional inclusion criteria for the present study. Given this set of core studies, a further criterion is that the proportion of blacks in the segregated control group must exceed 50%. This provision serves to conceptually tighten the notion of "segregation", and insures that the proportion of control group non-blacks in some studies will not approach the experimental group non-black proportions which are represented in others. The studies by Carrigan (1969) and Thompson & Smidchens (1979) were excluded from the analysis by this criterion.

The second part of Table 1 provides the guidelines for including the various segregated - desegregated comparisons which are contained within the 17 selected studies. The first restriction is that the Ns for both segregated and desegregated pre- and post-tests must be at least 10. This sets at least a moderate lower bound on the reliability of the estimates of sample means and standard deviations, as the precision of such estimates increases with sample size. Very small samples occasionally yield standard deviations which are only a fraction of the population value, and thereby are capable of producing highly misleading effect size estimates. A second inclusionary restriction on the particular comparisons concerns segregated control groups exposed to "enriched" or other novel types of curricula. Such control groups are not used because the resultant effect size estimates inversely reflect the efficacy of the particular special treatment employed in the "control" group. Such a situation fails to produce an acceptable test of the effects of desegregation on

black achievement.

Table 1

Inclusion Criteria

A. Criteria for inclusion of studies:

1. Study must be included in NIE core list.

2. Segregated control group must be over 50% black.

B. Criteria for inclusion of comparisons within studies:

1. *Ns* must be larger than 10 for both segregated and desegregated conditions.

2. Segregated control group must not receive any special treatments which extend beyond the typical classroom experience (e.g. "enriched" control classes are excluded).

3. Dependent variable must consist of a verbal, math, or "other" (e.g. science, social studies) achievement or ability test which corresponds to a major content area (excluded are IQ tests and "work study skills" tests).

4. Pretests and posttests must measure an identical construct.

5. Either:

a. Posttest standard deviations (or reliable estimates from national norms or a comparable study), along with pretest to posttest mean differences for segregated and for desegregated conditions, must be present; or

b. An ANCOVA table (with pretest differences as a covariate) which reports a *t* or an *F* value for segregated vs. desegregated posttest score differences must be present.

As indicated earlier, standardized achievement and ability tests of specialized content areas (e.g. social studies, science), as well as verbal and mathematical achievement, were included in the analysis. IQ comparisons were eliminated on the grounds that, in theory, a student's level of intelligence should not be especially sensitive to classroom experiences. Additionally, tests of "work study skills" were excluded because they do not correspond to any major academic content area. A further restriction noted in Table 1 is that the pretest and posttest had to measure an identical construct (e.g. "vocabulary", "arithmetic concepts"). Usually, this meant use of the same standardized tests (e.g. IOWA, Stanford, etc. - corresponding to the appropriate grade levels) for both the pretest and the posttest. However, cases in which the pretest and posttest differed, but nonetheless assessed the same construct, were also included, with the pretest means being adjusted to correspond to the posttest scale.

As noted in a preceding section, in studies of school desegregation researchers are rarely able to assign children randomly to experimental and control conditions. The selection effects that occur sometimes result in higher test score means and larger standard deviations in experimental than in control group prior to the onset of desegregated schooling. Therefore, it is important to attempt to correct post-measured differences so that they do not simply reflect the initial inequivalence of the comparison groups, but instead, reflect the effect of desegregated schooling.

In order to arrive at pretest-adjusted estimates of effect size, it is necessary to possess the following information: (1) an estimate of differential experimental vs. control group pretest/posttest gain scores; and (2) an estimate of the population standard deviation. Thus, the final criterion for inclusion listed in Table 1 is the presence of these two pieces of information. These numbers typically were furnished in the form of tables containing pretest and posttest means and standard deviations for both segregated and desegregated groups. Analysis of covariance summary tables (with pretest differences as a covariate) provided an acceptable alternative source of such information. Finally, in the absence of the above sources of information, a comparison could still be included if the pretest and posttest means were reported and if the standard deviation could be estimated from either national norms or from a comparable study using the same test for the same grade level.

Computation of Effect Size

The calculation of effect size estimates for the included comparisons was achieved via the following formula:

$$ES_1 = \frac{\bar{X}_{E(post)} - \bar{X}_{C(post)}}{\sqrt{\frac{(N_E-1)S^2_{E(post)} + (N_C-1)S^2_{C(post)}}{N_E + N_C-2}}} - \frac{\bar{X}_{E(pre)} - \bar{X}_{C(pre)}}{\sqrt{\frac{(N_E-1)S^2_{E(pre)} + (N_C-1)S^2_{C(pre)}}{N_E + N_C-2}}}$$

E=Experimental (Desegregated) Group

C=Control (Non Desegregated) Group

Effect size is defined here as the posttest desegregated vs. segregated difference in means (as expressed in pooled posttest

standard units) minus the pretest desegregated vs. segregated difference in means (as expressed in pooled pretest standard units). For the estimation of population pretest and posttest standard deviations, a pooled figure is used (in preference to Glass' recommendation of using only the control group standard eviation) in order to increase the reliability of such estimates.

Two points argue for the soundness of this procedure. First, the pretest control group standard deviations tend to be the smallest of the four sets of standard deviations (Experimental and Control pretest and posttest S.D.'s). Consequently, reliance on it for estimation of the pretest effect size that is to be subtracted from the posttest effect size will exaggerate the correction for pretest inequivalence of groups and thereby reduce the apparent effect of the treatment (desegregation) by too large a margin. Thus, a more reasonable procedure is one that employs an estimate based on a broader array of cases (Hunter et al, 1982). Adding to the soundness of using a population estimate based on a pooled figure is the fact that preliminary tests indicated that among the NIE core studies, no overall significant difference was present between the standard deviations of the desegregated and segregated groups at either the time of the pretest or the posttest.

Fan-Spread. It is important to note that the present effect size estimation procedure eliminates any interpretative problems stemming from the "fan-spread hypothesis". According to the fan-spread notion, a widening of the difference between group means over time will be accompanied by an increase in the within group

standard deviations. This implies that the difference between two group means may grow over time in the absence of any increment in the correlation between the treatment and the dependent variable (Kenny, 1975). The effect size formula used in this study, by separately standardizing the difference between means at times T_1 and T_2, permits a determination of the extent to which desegregation is associated with improvement in academic achievement over and above mere fan spreading. The computational procedure is identical to that used by Armor (1983) for those cases in which he judges fan-spread to be present. In other cases, however, a difference arises, in that Armor pools the four estimates of standard deviation in instances in which he judges that fan-spread does not exist.

Armor's procedure contains two problems. First, fan-spread is a matter of degree. What criteria should be used to make a dichotomous judgment of "present" or "absent" and how can such a dichotomous decision be justified? A statistical test of whether standard deviations differ in a particular instance is not a satisfactory criteria, in that it sensibly could be argued that correction should also be made when differences fall just short, or somewhat short, etc., of statistical significance.

A second problem is that Armor's procedure may systematically place undue weight on pretest differences. If it assumed that fan-spread effects do not occur, (or do not all of the time), and further that the distribution of pretest vs. posttest standard deviation differences is associated with a certain degree of sampling variance (which is particularly likely here due to small

sample sizes), then sampling error alone will produce a set of instances in which the pretest standard deviation is below the posttest standard deviation. This suggests that Armor's procedure may be susceptible to a bias in which only pretest standard deviations that happen to be low will be used to specifically scale pretest mean differences, while those that are higher (relative to the posttest standard deviation) will be averaged in with the posttest estimates. The net result is that pretest differences may be given a disproportionately high weighting across cases. Because the desegregated group usually shows a higher pretest mean than the segregated control group, Armor's procedure consequently can be expected to produce a lower overall estimate of effect size than the formula that I will be using.

In order to assess the extent to which a consideration of fan-spreading, however, is important in accounting for the results of the current sample of desegregation studies, effect size estimates were also calculated by using an alternative formula:

$$ES_2 = \frac{(\bar{X}_{E(post)} - \bar{X}_{E(pre)}) - (\bar{X}_{C(post)} - \bar{X}_{C(pre)})}{\sqrt{\frac{(N_E-1)S^2_{E(post)} + (N_C-1)S^2_{C(post)}}{N_E + N_C-2}}}$$

E=Experimental (Desegregated) Group

C=Control (Non Desegregated) Group

In this formula, the desegregation vs. segregation pre-post gain score difference is divided by an estimate of standard deviation that is based on the pooled posttest figures. If the pretest standard deviations tend to be low relative to those of

the posttest, and if the desegregation group tends to possess a higher mean than the control group at the time of the pretest (as is the case when the fan-spread hypothesis holds), then this formula should produce larger estimates of effect size than should the first formula. This is true because the typical pretest advantage for the desegregated students, which is subtracted from the standardized posttest difference, will be weighted more heavily in determining effect size estimates.

Effect size estimates based on analysis of covariance. For cases that only reported an ANCOVA (Analysis of Covariance) summary table, in which pretest scores served as the covariate, the following transformation procedure was used to estimate the effect size:

$$ES = t \frac{2}{\sqrt{N}} (.633)$$

where N is the combined sample size. Multiplying by .633 serves to correct for the fact that the variance of change scores tends to be lower than the variance of raw sample scores: ($S^2_{change} = 2S^2(1-r)$ as reported by Armor), with the difference being greatest for cases involving high pretest-posttest reliabilities. For the present purposes, a fairly high reliability estimate (r=.8) was assumed, which algebraically leads to the modification of effect size noted above.

Sample size. Some experts (e.g. Hunter, et al.) argue that a summary statistic of the effect sizes computed for the sample of studies (viz. mean effect size) should be weighted by the sample size of each study. Though there often may be good reasons to

adopt this procedure, especially when summarizing experimental studies, for several reasons, it will not be used here. In experimental research, the manipulations are designed to correspond to a theoretical variable. Researchers almost routinely use manipulation checks to assess whether or not the independent variable theoretically postulated to affect the dependent measure has in fact been manipulated by the experimental operations that were employed, and if so, to assess whether it was manipulated "strongly enough". If, in a particular study, the manipulation check failed to confirm appropriate variation of the independent variable, no sensible scientist would want to include the study in the meta-analysis.

In contrast, as I have argued above, it is not clear what, if any, theoretical variable corresponds to or is conceptually linked to a change in the ratio of black and white children in a classroom (or school) and consequently, might be responsible for black achievement gains. Indeed, as indicated later in this paper, my own research seriously impugns any positive role for the one theoretical process postulated in the past to cause academic gains for minority students. Not knowing what underlying theoretical variable is relevant to academic gains for blacks, it makes perfect sense that such manipulation checks simply are not found in desegregation research. Consequently, one cannot know whether or not in any particular study the desegregated groups were exposed to the "key ingredients". If a study with a very large sample fails to contain these ingredients (or contains other features which produce losses in black achievement), and if this

study outcome were weighted by its sample size, it might more than counterbalance the effects of other studies, which, with smaller samples, produced positive effects. (In this regard, it is noteworthy that sample sizes among studies in the NIE core set vary by a margin of fifty to one). Stating this another way, extraneous factors related to sample size, which may or may not be causal, may be correlated with effect size.

Anticipating the results, analyses show that: (1) sample size is indeed negatively correlated with effect size (r=-.404) and (2) the observed variation among effect sizes exceeds that to be expected from sampling error, suggesting that moderator variables are in fact operating. Taken together, these considerations argue strongly for the decision to weight study outcomes equally, rather than by sample size.

Correction for unreliability. In the current analysis, each effect size estimate was corrected for unreliability (following the procedures of Hunter et al., 1982). Measurement unreliability has the effect of artificially inflating the variability of scores, thereby leading to larger standard deviations and, hence, lower absolute values of effect size estimates. The unreliability correction procedure advanced by Hunter, et al., divides the estimated effect size value by the square root of the reliability coefficient of the dependent measure. In some of the cases comprising the NIE core studies, reliability coefficients were either reported directly or were readily available from national norms. For the remainder, a conservatively high reliability estimate of .95 was automatically assumed for each test. The net

result of correcting for unreliability was to increase the absolute value of the particular effect size estimate by about 1.5% to 3%.

Results

The results of the meta-analysis are summarized in Table 2. For each study, a mean was calculated (when possible) for each of the three types of dependent variable categories (i.e., verbal, math, and "other"). Next to each mean, in parentheses, is the number of different tests that were averaged in arriving at the figure.

Using formula (1), the overall effect size is +.192 (see bottom of column 1, Table 2). This estimate weights results within each study equally and weights each study equally. The fact that formula (2) gives an outcome of +.184, which is essentially equivalent to that obtained with formula (1), confirms the view, presented earlier, that fan-spread is not a problem in these data.

For purposes of comparison, the effect size computations of Armor (1983), Stephan (1983), and Wortman (1983) are reported in the adjacent columns of Table 2 (columns 3, 4, and 5). Table 3 summarizes the findings of all four researchers, reporting their mean effect sizes, separately for verbal and math tests, for each study. Pooling the outcomes across researchers and studies, the effect size of +.164 for verbal tests is significant ($t=2.34$, $p < .05$), as is the pooled verbal and math effect size of +.119 ($t=2.63$, $p < .05$). The effects of desegregation on mathematics tests is smaller than that found on verbal tests (though not

Table 2

Effect Size Estimates

Study	Year	Miller (#1)[a] Verbal	Miller (#1)[a] Math	Miller (#1)[a] Other	Miller (#2)[a] Verbal	Miller (#2)[a] Math	Miller (#2)[a] Other	Armor Verbal	Armor Math	Stephan Verbal	Stephan Math	Wortman Verbal	Wortman Math
Anderson	1966	+.733 (3)[b]	+.669 (1)	---	.764(1)	+.671(1)	---	+.89 (1)	+.54 (1)	+.42	+.24	+.95 (1)	+.51 (1)
Beker	1967	+.400 (7)	-.203 (3)	+1.128(1)	+.265(7)	-.103(3)	+.942(1)	+.25 (2)	-.16 (2)	+.125	-.24	+.095(2)	+.285 (2)
Bosman	1973	+.061 (1)	-.160 (2)	---	+.033(1)	-.375(2)	---	-.26 (2)	-.21 (2)	+.125	+.055	+.02 (1)	-.06 (1)
Carrigan	1969	---	---	---	---	---	---	-.060(6)	---	-.04	---	-.04 (6)	---
Clark[c]	1971	+.039 (2)	-.132 (1)	---	d	d	---	-.01 (1)	+.12 (1)	+.08	-.24	e	e
Evans	1973	+.040 (2)	+.084 (2)	---	d	d	---	+.015(2)	+.07 (2)	+.02	+.03	e	e
Iwanicki and Gable[c]	1978	-.051 (3)	---	---	d	---	---	0.0 (3)	---	---	----	e	e
Klein[c]	1967	+.159 (1)	+.333 (1)	+.313(1)	d	d	d	0.0 (1)	-.08 (1)	+.21	+.33	e	e
Laird and Weeks	1966	+.257 (3)	+.074 (3)	---	+.341 (3)	+.127(3)	---	-.123	-.06 (3)	+.13	+.063	e	e
Rentsch	1967	+.617 (3)	+.562 (3)	---	d	d	---	---	---	+.22	-.053	+.497(3)	-.110 (3)
Savage	1971	-.080 (3)	-.091 (1)	+.144 (3)	-.077 (3)	-.014(1)	+.030(3)	+.15 (1)	-.08 (1)	+.06	-.04	+.14 (1)	-.05 (1)
Sheehan and Marcus[c]	1978	-.188 (2)	-.163 (1)	---	d	d	---	-.16 (1)	+.21 (1)	-.07	-.08	-.16 (1)	-.16 (1)
Sloan[c]	1968	+.091 (2)	+.291 (1)	---	d	d	---	+.27 (1)	+.47 (1)	+.19	+.22	e	e
Smith	1971	-.077 (3)	+.130 (1)	---	-.101 (1)	+.256(1)	---	-.06 (1)	+.13 (1)	-.01	+.02	-.05 (1)	+.10 (1)
Syracuse	1966	+.691 (1)	---	---	+.691 (1)	---	---	+.375(2)	---	+.75	---	e	---
Thompson and Smiddiens	1979	---	---	---	---	---	---	---	---	-.15	+.04	e	e
Van Every	1969	-.166 (2)	+.543 (1)	---	-.217 (2)	+.559(1)	---	-.46 (1)	+.51 (1)	-.12	+.14	-.44 (1)	+.53 (1)
Wallberg	1971	-.049 (3)	---	---	-.048 (3)	---	---	-.10 (4)	---	+.055	-.02	+.012(4)	?
Zlep	1971	+.671 (2)	-.157 (1)	---	+.667 (2)	-.157(1)	---	+.53 (1)	-.17 (1)	+.066	-.15	+.065(1)	-.15 (1)
		N = 17	N = 14	N = 3									
		$\bar{X}$ = +.185*	$\bar{X}$ = +.127	$\bar{X}$ =+.528									
		SD = .117	SD = .304	SD = .526									

N = 14

$\bar{X}$ = +.192**

SD = .338

Notes:

a. See text for formulas #1 and #2.

b. Numbers in parentheses are the number of effect size comparisons.

c. Uses estimates based on ANCOVA.

d. Estimates from formulas #1 and #2 are identical due to use of ANCOVAS

e. Not pretest adjusted.

* $p < .05$; ** $p < .01$.

Table 3
Mean Effect Size Estimates

Study	Verbal	Math
Anderson	+ .75	+ .49
Beker	+ .22	- .08
Bowman	- .01	- .09
Carrigan	[- .049	---]
Clark	+ .04	- .16
Evans	+ .03	+ .06
Iwan. & Gable	- .03	---
Klein	+ .13	+ .19
Laird & Weeks	+ .24	+ .03
Rentsch	+ .44	+ .13
Savage	+ .07	- .07
Sheeh. and Marcus	- .14	- .15
Stone	+ .18	+ .33
Smith	- .05	+ .10
Syracuse	+ .61	---
Thompson & Smid	[- .15	+ .04]
Van Every	- .30	+ .43
Walberg	- .02	- .02
Zdep	+ .63	- 16

			Combined V&M
N	17	15	32
$\bar{X}$	.164 [b]	.069	.119 [c]
SD	.289	.211	.256

[a] Entries combine the computations of Miller (#1), with those of Armor, Stephan, and Wortman. Excludes Carrigan, Thompson and Smidchens.

[b] $t_{(16)} = 2.34$, $P < .05$

[c] $t_{(31)} = 2.63$, $P < .05$

significantly so) and when tested separately, does not yield a significant effect size (see columns 1 and 2, and see Table 3).

Sources of Disparity in the Effect Size Estimates for Individual Studies

Comparison of my own effect size computations with those of Armor, Stephan, and Wortman for each study reveal that they agree fairly well; the correlations, using estimates based on formula (1) are +.87, +.77 and +.78 with Armor, Stephan, and Wortman respectively.

The correlations were computed by treating the mean verbal effect size per study and the mean math effect size per study as separate entries. The fact that the verbal and math effect size estimates are not based on independent samples is irrelevant for this computation in that it seeks to assess the comparability of effect size computations performed by independent investigators. There is little reason to think that computations performed within a study are less independent than those between studies. Despite the high correlation between estimates, the fact that these correlations are less than perfect, as well as the fact that inspection of effect sizes across the rows of Table 2 reveals variation, makes it clear that computational differences exist.

The following paragraphs, on a case by case basis, examine all instances in which my estimates differed from the mean estimate of Armor, Stephan, and Wortman by more than .1 of a standard deviation.

Anderson (Math)

My estimate is slightly higher (+.669) than those of Armor

(+.54) and Wortman (+.53), mainly as a result of discrepancy between the mean of the raw pretest segregated math scores contained in Table 26 (45.093, p. 138) and the mean he presents in his pretest summary table (43.82, p. 144). I used the mean of the raw scores, which led to a higher effect size estimate due to the inclusion of a larger segregated group pretest figure.

Beker (Verbal)

The major reason for my higher estimate seems to be my inclusion of a wider array of tests (spelling, word meaning, language, and vocabulary) which demonstrated larger positive effects than did paragraph meaning. Wortman's estimate is additionally lower due to his exclusive use of the "refused transfer" controls instead of the "requested transfer" group.

Klein (Math)

My estimate for math agrees with that of Stephan (+.33), but is substantially higher than Armor's (-.08). The reason for the discrepancy is that I used only the "random" control group, while Armor used only the "matched" control group. The matched controls were excluded from the present analysis because the corresponding ANCOVA summary table mixes the data for the segregated and desegregated blacks along with that of the white students.

Syracuse (Verbal)

The present figure for the Syracuse report (+.691), while relatively close to Stephan's estimate (+.75), is much higher than Armor's (+.375). The reason is that Armor includes a second comparison (which I excluded because of missing standard deviations) in which the effect size was essentially zero.

Van Every (Verbal and Math)

My estimate for verbal achievement (-.166) is somewhat less negative than the estimates of Armor (-.46) and of Wortman (-.44). This is because they only consider Reading (which I estimated at -.468), while I additionally included Language Arts (+.137).

My math estimate is nearly identical to those of Armor and Wortman, and differs significantly only from Stephan's figure. Stephan's lower estimate most likely stems from his use of Glassian formulas, in conjunction with his correction procedure for the amount of time elapsing between the pretest and the posttest.

Walberg (General Note)

Due to problems in the legibility of my copy of this report, I was unable to calculate a verbal effect size estimate for the 10-12th grade group, as well as any estimates for math achievement.

Sources of Disparity in Overall Effect Size Estimates

Among the three NIE panel member's computed effect size estimates, Armor's overall effect size estimate of +.077 is most discrepant from my own. Consequently, his computations were chosen as a basis for estimating sources of discrepancy.

Table 4 presents an analysis of the disparity. It shows that correction for unreliability in the dependent measures is not a major contributor to my higher estimate. In part, this is due to the fact that conservatively high reliability estimates (viz .95) were assumed for the studies for which no reliability was reported. Reliability estimates provided by test publishers do

Table 4

Analysis of Discrepancy between Effect Size Estimates of Armor and Miller (#1)[a]

Source	Contributions
Inclusion of Reliability Correction	+ .005
Inclusion of Rentsch + "Other" Category Data	+ .062
Averaging in of Extra Tests Excluded by Armor	+ .002
Calculational Differences on same Non-Ancova cases	+ .006
Calculational Differences on cases where I estimated from Ancova	- .006
Different comparison Groups used in same study (Klein)	+ .0172
Armor's Inclusion of Carrigan Study	+ .005
Cases within studies included only by Armor	+ .022
Total:	+ .1132
(Miller + .192) - (Armor + .077) =	+ .1150
Unaccounted difference =	+ .0018

Note:

a. Table entries are based on overall means of Miller's Verbal, Math, and "Other" tests.

not report separate reliability estimates for blacks, but were they available, they are likely to be lower than those reported for whites. In sum, a less conservative and more realistic correction for unreliability would yield a larger, more positive overall effect size estimate.

The category responsible for the largest portion of the difference (over 50%) is the inclusion of the Rentsch study (also included by Stephan and Wortman) and the inclusion of results on achievement tests on content other than verbal skills and mathematics. It is worth noting that although only three studies report such results, the mean effect size (and its standard deviation) is substantially larger than that of effect sizes based on verbal and mathematics tests.

Moderator Variables

Ordinarily, with such a small set of studies, it is hard to justify a search for variables that explain the relation between the independent (school desegregation) and dependent (academic achievement) variables. A simple set of computations, however, can suggest whether such a search will be fruitful. The variance of the effect sizes over the sample studies can be computed and corrected for sampling error. If the effect sizes are really identical and vary only because of sampling error (i.e., they are simply random deviations from the true mean value), then the true variance of the effect sizes would be zero. Hunter, et al., provide formulas for computing the variance of an array of effect sizes, corrected for sampling error. When sampling variability (σ^2_{error}) is removed from the computed variance among obtained

effect sizes (σ^2_{ES}) there should be no residual (viz. $\sigma^2_{ES} - \sigma^2_{error} = 0$) if, in fact, the effect size is really the same across studies. If, on the other hand, the residual variation is large, especially if large in comparison to the mean value, a search for moderator variables should be made.

In the present case, the effect sizes for verbal achievement tests were used to assess this issue. When sampling variablity is removed, the residual variance does not approximate zero. This is true irrespective of whether one uses an estimate of the average effect size that is unweighted by sample size

$$(\sigma^2_{ES} = .079;\ \sigma^2_{error} = .012)$$

or weighted by sample size

$$(\sigma^2_{ES} = .049;\ \sigma^2_{error} = .012).$$

These results show that 82% or 67% of the variance in the computed effect size scores (unweighted or weighted by sample size respectively) is unexplained by sampling error.

$$\text{Explained Variance} = 1 - \frac{\sigma^2_{error}}{\sigma^2_{ES} - \sigma^2_{error}}$$

These results argue strongly that variation among study characteristics and not mere sampling fluctuation is responsible for the observed variation in the computed effect sizes.

Given these results, three potential moderator variables were examined: year of study, region (North vs. South), and percentage of black students in the desegregated class. Prior to computing the correlation between effect size and each potential moderator variable, I averaged my own effect size estimates with those of Armor, Stephan, and Wortman, separately for verbal and math

achievement. Pooling gives a more stable estimate. Although earlier in the chapter I argued that the different content domains of academic performance should be considered indices of a common underlying construct, separate treatment of verbal and math effects is justified by the low correlation between these two effect sizes estimates within each study ($r = +.29$; $r^2 = +.084$; $df = 12$; $p > .05$), and the fact that Stephan provides a theoretical rationale for different outcomes on verbal and math tests. When the verbal and math effect sizes of Armor, Stephan, and Wortman are pooled with my own, the correlation between them is even smaller ($r = +.15$; $r^2 = +.023$; $df = 12$; $p > .05$).

Since effect size estimates contain sampling error, correlations will be attenuated in the same fashion that correlations ordinarily are attenuated by measurement error. Therefore, the correlation between effect size and each moderator variable was adjusted as follows:

$$\text{Rel. of ES} = \frac{\sigma^2_{ES} - \sigma^2_{error}}{\sigma^2} = \frac{.079 - .012}{.079}$$

$$\text{Corrected Correlation} = \frac{r_{(ES,X)}}{\sqrt{\text{Rel.}_{ES}}}$$

Interestingly both verbal and math effect size estimates correlate negatively with year of study ($r_V = -.563$ and $r_M = -.560$, $p < .05$ uncorrected respectively; $r_V = -.611$, $r_M = -.608$ corrected). Region is unassociated with effect size (point biserial: $r_V = +.121$; $r_M = +.025$, north higher, $p > .05$).

There is some suggestion, however, that percentage of blacks in the classroom is important and that it has different effects on

verbal and math achievement. The correlation between percentage of black students in the class and verbal effect size is -.344 (corrected for reliability), indicating that the fewer same-race peers a black child finds in his or her desegregated classroom, the greater the ensuing improvement in verbal achievement. (When year of study is partialed out, the correlation increases to -.42). In contrast, no such effect is found for math achievement; in fact, the correlation between percentage black and math achievement, though not significant, is opposite in sign (+.181). When year of study is partialled out, the difference between these correlations approaches significance ($p<.05$, one-tailed).

These results provide some support for Stephan's (1983) interpretation of his own computed effect size differences for verbal and math achievement, showing desegregation to produce essentially no benefit for the latter. He interprets the gain in black verbal achievement that is found with desegregated schooling to be a consequence of increased exposure to white speech style, syntax, grammer, etc. If this interpretation has merit, it makes sense that percentage of blacks in the classroom should be inversely related to such gains. The fewer the number of other blacks in the classroom, the more likely it is that the desegregated black child must interact with white children and the less likely it is that he or she would find a within-race peer support group in which black speech is practiced and reinforced.

Correction of Effect Size Estimates for "Overall School Improvement"

The analyses presented above examine the achievement gains of desegregated black children but ignore changes among their white classmates. It is imporant to examine them, however, because when both groups gain (or lose) it suggests that it is not desegregation per se that is responsible for the effect, but instead, some other factor that has affected the school or school district as a whole, thereby improving the academic performance of all of its students. Such factors might be: influx of new funding; improved curriculum materials; a new principal; renewed teacher enthusiasm; increased emphasis on preparation for state-mandated testing; or whatever.

Those sympathetic to the idea of desegregation might contend that when school changes such as those cited above appear hand in hand with desegregation, they should not be viewed as confounding effects, that is, as factors other than desegregated schooling that explain the observed minority gains. Instead, they should be thought of as natural covariates of desegregation, that is, as part of the meaning of the term. In other words, according to this line of thought, whenever one desegregates a school or school district these simultaneous changes (whatever they are, and however unspecified they must remain) can be expected to co-occur with the change in the ratio of black and white students. And as long as they regularly or naturally co-occur with desegregation, their academic benefits to minority children can be attributed to desegregation. In this view, if whites gain along with blacks,

all the better.

There are two problems with this line of thought. One lies in the validity of the assumption that these school changes can be expected to co-occur routinely with desegregation in the future (or in other unsampled districts). For instance, today, in an era of minimal availability of increased state and federal funding for schools, some of these mediating factors (e.g., new or improved curriculum and/or text materials, or lower pupil-teacher ratios) may no longer be readily available to desegregating districts. Similarly, 15 years ago teachers and principals may well have been more inclined to expect positive outcomes as a consequence of desegregation than they do today. Such expectancies have often been found to be self-fulfilling for one reason or another. If present then, but not today, outcomes would again differ depending on whether one included or excluded such factors in one's definition and implementation of desegregation. The strong negative correlations reported above between year of study and positivity of both verbal and math effect size estimates argues strongly that one cannot rely routinely on the natural occurrence of these beneficial ingredients.

A second problem lies in one's definition of academic benefit. Some scholars argue that benefit should be defined in an absolute sense. If desegregation produces academic gains for blacks, and does not produce losses for whites, it is beneficial. In this view, it does not matter if the gains of white children equal or exceed those of blacks. An alternate view focuses instead on the closing of the academic achievement gap.

Consequently, it defines desegregation as beneficial only if the gains of black children exceed those of whites.

Three studies in the NIE core set, Beker (1967), Clark (1971), and Laird and Weeks (1966), provide data that permits analysis of the effects of desegregation on white as well as black children. All seven available cases of the mean verbal, math, or "other test" effect size per study can be compared by using the following formula:

$$\left(\begin{matrix}\text{Desegregated} \\ \text{blacks}\end{matrix} : \frac{\bar{X}\ \text{post} - \bar{X}\ \text{pre}}{\text{pooled pre + post SD}}\right) - \left(\begin{matrix}\text{Receiving School} \\ \text{whites}\end{matrix} : \frac{\bar{X}\ \text{post} - \bar{X}\ \text{pre}}{\text{pooled pre + post SD}}\right)$$

The resulting difference in effect sizes is -.379, (N=7, $p>.05$, S.D.=.894). Although not significant with only seven cases, the direction of effect shows that the gains of white children in the receiving schools of these studies substantially exceeded those of black children, which were roughly of the same positive magnitude as the gains found for the entire sample of blacks. That is, the mean effect size for blacks in these three studies (weighting tests equally) was +.15, (compared to the entire sample effect size of +.192), whereas the effect size for whites was +.52. In other words, the achievement gains of white children in these three studies were more than three times as large as those of their black classmates.

In summary, on the basis of this extremely small subsample, it appears that black gains relative to white gains were small. In terms of the preceding discussion, these data suggest that the observed gains of desegregated black children are not attributable

it is not attributable to desegregation per se, but instead, to other school or district factors that accompany its implementation.

Factors Affecting Academic Outcomes in Desegregated Settings

As stated above, there is little good theoretical understanding of how desegregated schooling might improve the academic performance of minority children. Much past theorizing has not withstood the test of data. The next section briefly discusses an array of factors, some of which were thought in the past to be relevant and some of which continue to appear important.

Anxiety and threat. The fact that high anxiety impairs performance on complex or difficult tasks fits with common sense and is one of the better established findings of psychology. In his review of variables that affect black performance on cognitive tasks Katz (1968) summarized substantial evidence showing impairment when performing under the scrutiny of higher status whites. The administration of standardized achievement tests to black students by a white teacher in a white dominated setting, such as a desegregated classroom, structurally parallels the situations studied and cited by Katz as impairing black performance. The fact that standardized achievement tests are administered with time limits acts to further raise anxiety. Some evidence suggests that one-way busing of blacks to white receiving schools will increase their anxiety in general, at least during the initial phases of desegregation (e.g., Gerard & Miller, 1975).

Mussen (1953) found that black children perceive more hostility or threat in their environment than do whites. Baughman (1971) interprets the heightened level of worry and anxiety that black children attribute to their characters when asked to make up stories as confirming Mussen's results.

Taken together, such data implies that measured black performance is likely to be an underestimate of true mastery; it implies that the obtained effect sizes for black academic achievement do not reflect true level of achievement. But if adult black intellectual activity is performed in a white world, aren't such depressed scores in fact legitimate scores? Perhaps, but in work settings performance is rarely under the constant scrutiny of a white supervisor.

Self-concepts and aspirations. In the social science statement appended to Brown, scholars argued that segregated schooling lowered the self-concept of the minority child and that this in turn produced a sense of defeatism, self-doubt, and lack of aspiration that interfered with effective learning. Although the argument appears credible, it has not withstood empirical analysis. Not only has the interpretation of Clark's (1937) original doll preference data on which the argument was based been questioned (Brand, Ruiz, & Padilla, 1974; Banks, 1976), but recent reviews of self-esteem research that employs direct self-report measures consistently show either higher levels of self-esteem among black children than among white children or no consistent effects (Epps, 1979, Porter & Washington, 1979, St. John, 1975, Stephan, 1978, Wylie, 1979). Furthermore, if school desegregation

does affect the self-esteem of black children, its effects, at least initially, are more likely adverse than positive (Porter & Washington, 1979).

Measures of aspirations present a similar picture. Black children in segregated schools typically report higher aspirations than do white students (Epps, 1975; Proshensky & Newton, 1968; Weinberg, 1975). And black adults seem to value education more strongly than do whites (Wilson, 1970). The effect of desegregated schooling on the motivation of black students remains unclear, some studies showing higher black aspirations in desegregated schools (Curtis, 1968; DeBord, Griffen, & Clark, 1977; Fisher, 1971; Knapp & Hammer, 1971, Reniston, 1973), others showing an opposite effect (St. John, 1966; White & Knight, 1973; Wilson, 1959), and still others showing little difference between black children who attend segregated or desegregated schools (Curtis, 1968; Falk, 1978; Hall & Wiant, 1973). Two points must be made with respect to this issue. First, most experts today would agree that level of aspiration per se is not as meaningful or important an indicator of a healthy personality as is a level of aspiration that is in line with one's level of performance and one's obtained outcomes. Second, the nature or design of these studies does not allow causal interpretation of whatever differences are found.

Finally, although the theorizing of social scientists at the time of Brown allowed for circular feedback loops (or bi-directional or reciprocal causation) between self-esteem, motivation and aspiration, intergroup acceptance, and academic

performance, their arguments clearly emphasized a causal pattern in which personality variables (self-concept and achievement motivation) caused subsequent changes in academic performance. If there is any preponderent direction of causal effect, researchers today would emphasize the impact of school outcomes (academic performance and achievement) in forming personality or creating changes in it, rather than a causal pattern in which changes in personality cause subsequent shifts in performance (Gottfredson, 1980; Miller, 1982; Rubin, Maruyama, & Kingsly, 1979; Scheirer & Kraut, 1979).

Peer comparison. When black children attend desegregated rather than segregated schools, social comparison between their own academic performance and that of white students will reveal disparities that might be expected to lower their academic self-concepts and lead to self-definitions of poor ability on these tasks. This in turn should act to lower performance. If such effects occur, they should be greater at higher grade levels in that on the average the academic disparities between black and white students increase as they progress through school.

On the other hand, other data suggests that black children primarily compare themselves to other black children (Baughman, 1971). To the extent that the desegregation plan provides enough black children in each class to form the basis for a within-race comparison group, the debilitating effects of comparison with white children should be lessened. Perhaps in part to cope with such invidious comparison, black children develop defense mechanisms for themselves and their friends that shield them from

evaluations that are threatening. Students know who is smart and who is not (Lippit & Gold, 1959; Hoffman & Cohen, 1972). Differences in opportunity to perform, when coupled with a narrow range of valued abilities, act to create widely shared perceptions of competence (Simpson, 1981; Rosenholtz & Rosenholtz, 1981). However, children, like the rest of us, are self-protective and adaptive. They find ways to ignore self-disparaging comparisons and, as evidence on black children's self-esteem and aspirations shows, if anything, in their self-reports these children show high levels of self-regard and expectation. Whether or not these high levels are "defensively high" as suggested by Entwisle & Hayduk, (1982) and Miller, (1982), and reflect a negative consequence of peer comparison remains unclear.

Expectations. As indicated above, expectations often create self-fulfilling cycles. Expectations to perform poorly cause behavior that subsequently confirms the expectation. But expectations are intimately linked to actual behavior. Rehearsal of academic information and content improves performance on subsequent testing of the mastery of this information. It is the better student who volunteers the answer when the teacher calls for a response, who leads the discussion in peer tutoring or small work group exercises, and who the teacher routinely gives more opportunities to respond (Good, 1970). Thus, it is the better student who gets the benefit of overt rehearsal at the expense of less capable peers, thereby further improving the performance of the better student. The social dominance of whites when in interaction with blacks is well documented. Even when the

resources and knowledge brought to the problem by black and white children is equivalent, the white child will initiate verbal comments more often than the black and will dominate the interaction, with the black child taking a more subordinate role (Cohen, 1982). Apparently, generalized status differences are implicit in the distinction between races. Even when black students are primed with correct information that makes them a more superior source of knowledge than the white children, the generalized status difference between blacks and whites nevertheless results in continued verbal dominance by the white children (Cohen & Roper, 1972; Tammivaara, 1982).

Peer relations. Some social scientists believed that the peer environment of the desegregated school would be critical in producing academic gains (Coleman et al. 1965; Crain & Weissman, 1972; Pettigrew, 1969). This belief rested on the assumptions that (a) the student body of a desegregated receiving school is more likely than that of a segregated school to be of middle class family background; (b) middle class students are more strongly oriented toward achievement and thereby create a normative structure that emphasizes it; and (c) provided that the number of white students in the receiving school exceeds the number of incoming minority students, the latter group will adapt to the prevailing norm structure of the middle class whites. This argument, spelled out in detail by Katz (1964), rests on the additional assumption that minority children will be accepted or befriended by white children.

The latter assumption is at best, less true than one might

wish. Resegregation is common in desegregated classrooms, (e.g., Rogers & Miller, 1980; Rogers & Miller, 1981; Schofield, 1980) and when white children accept minority students it is a consequence of the minority students' good academic performance rather than a cause of it (Maruyama & Miller, 1979; Maruyama & Miller, 1983). Thus, it is not the peer system that provides a critical normative influence. Instead, as discussed in more detail below, it is provided by the teachers and administrators.

School effects. Recent research, Jenks et al. (1975) notwithstanding, shows that schools can exert powerful educational effects on students (Heyns, 1978) and differ in the extent to which they educate them (Edmonds, 1976). These effects are system or organization effects, produced in concert by principals, teachers, students, neighborhood, parents, and all having reciprocal influence on one another. This is not to argue that one cannot find, for instance, within-school differences among teachers both in their background and their approach to education, or differences among students. It startles no one when a low social class background is found to be related to a student's academic performance (Hauser, 1978). Nor does it elicit much more surprise to learn that the quality of teachers' education affects the academic outcomes of their pupils (Heim, 1970; Summers & Wolfe, 1977). More interesting, however, are the substantial differences in academic outcomes found among schools whose students are basically similar in social class background and/or race. Although some authors have argued that such school effects are small (e.g., Sewell, Haller, & Portes, 1969), the studies on

which such conclusions are based all use high school samples. By high school age, self-fulfilling characteristics of background, expectations, and scholastic outcomes have homogenized schools, not unexpectedly leaving them similar in their educational impact, and consequently, leaving the false impression that the type of school attended cannot make a difference. At earlier ages, however, the homogenization process is not complete. Interestingly, studies of elementary schools do show striking differences between schools.

Two recent studies dramatically illustrate the powerful differences among schools in their effects on students (Brookover, Beady, Flood, Schweitzer, Wisenbaker, 1979; Entwisle & Hayduk 1982). Both are very substantial in terms of their breadth and the array of measures they employ. The Brookover et al. study is based on data from over 11,000 students in the fourth and fifth grades in over 90 schools drawn by random from the entire State of Michigan. Among those, 30 are majority black schools. This exceeds the totals of students and schools in the entire array of the nineteen NIE sample desegregation studies by a margin of about 3 to 1. Entwisle and Hayduk (1982) studied approximately 1,500 children over a three-year period from first to third grade. Approximately one-third, respectively, attended a white middle class school, an integrated lower class school, and a black lower class school. Although much smaller in terms of the number of schools studied, this study measured an even broader array of variables than the Brookover et al. study and on each, took multiple (longitudinal) measurements on each child over the three-

year course of the study, thereby enabling study of the temporal changes on the measured variables. It is only with temporal spacing of repeated measures on the same child that one can begin to establish the causal connection between variables. Thus, the two studies differ substantially in the characteristics of their research designs. Nevertheless, as will be indicated below, their results converge in identifying key aspects of the process of education, as well as showing that schools can produce very different outcomes for children.

Teachers. Earlier work demonstrated that teachers exert powerful effects on minority student outcomes (Johnson, Gerard, & Miller, 1975; Fraser, 1981). When desegregated minority children are imbedded in the classes of prejudiced teachers their academic performance worsens, whereas in the classes of unprejudiced teachers, it improves (Johnson, Gerard, & Miller, 1975). Furthermore, these effects can be traced to clear differences in the way in which these two types of teachers conduct their classes and interact with minority students (Frazer, 1981). This conclusion is supported by Brookover et al. and by Entwisle and Hayduk. In some lower class black schools the teachers (and the principal) have given up on the students. They do not view their students as capable of learning, attributing their poor academic outcomes to their backgrounds and not demanding good and consistent work from them. It is important to emphasize, here, that it is not merely teacher's expectations that produce these effects, but instead, it is their behavior. In lower class black schools that produce poor academic outcomes, students are not

expected to perform up to grade level and demands requiring them to do so are not placed on them. When teachers judge their students to be incompetent they do not attempt to cover as much academic material (Beez, 1970).

Teachers in most lower class schools also fail to voice concrete achievement goals. Instead, these children are often reinforced for incorrect performance, hearing the teacher say, for instance, "good try" when the answer is very clearly wrong, or not receiving immediate re-instruction when their response is incorrect (Brophy & Good, 1970). Academic norms of high academic achievement are recognized in high achieving lower class black schools, whereas such norms and a commitment to academic mastery are missing in the low achieving schools. In the high achieving schools, teachers spend most of the day instructing their students, reinforcing them discriminantly rather than indiscriminantly. In these schools, teachers do not highly differentiate among students and, in the process, write off a large segment of them as unteachable.

Students. Although many factors may contribute to the greater sense of control over their outcomes in life seen in middle class as opposed to lower class children (Coleman et al. 1966), the schools they attend seem to contribute to this observed difference. The students in low achieving schools show a legitimate sense of futility. With reason, it is difficult for them to know what to expect. The messages they get confuse and demoralize them. The teacher says, "Good, you're trying hard"; "OK," but they receive C's and D's on their report card.

Consequently, their expectations are not responsively modified by their obtained grades. In contrast to a sense of mastery and control of their academic outcomes, these students feel the system is whimsical and "stacked against them." In contrast, children in high achieving middle class schools increasingly come to forecast their school outcomes accurately. Their expectations more closely correspond to the grades they receive, with most students predicting their marks correctly (Entwisle & Hayduk, 1982). Brookover et al. (1982) argue that a sense of control over school outcomes is one of the essential ingredients for high student achievement.

Implications of Academic Achievement Results in the Context of Educational Goals

What does one make of the moderate positive effect of desegregation on the academic achievement of black children? Although not a strong clarion for desegregation in its own right, it certainly is not a deterrent to the continuation of desegregation as a national policy. More important, however, is the fact that other valuable educational goals cannot be met without desegregated schooling. Although cognitive development and academic mastery are obviously appropriate educational goals, they are not the only ones. Despite some recent signs of increased interest in "fundamental" education, all school curricula to some degree attend to dimensions other than verbal and mathematical skills. Indeed, many components of the standard educational curriculum attend to dimensions that have little or no

direct relevance to cognitive mastery, e.g., physical education; music, art, and aesthetic development; mechanical, shop, and home skills; industrial, business, and other vocational training; etc.

In some sense all agree that schools must prepare children to function effectively in their adult life. Thus, some view with despair the tracking of students within performance levels and in qualitatively different academic programs because it functions to prepare students for occupational and social roles that reflect their socioeconomic origins (Bowles & Gintis, 1976); and students within the different tracks do display attitudes and patterns of interpersonal behavior that are complementary to these future roles (Oakes, 1982).

Similarly, few would argue against the view that interpersonal skills are relevant to accomplishment and success in adulthood. In a multi-ethnic society, constructive modes of interethnic interaction, as well as interethnic acceptance and trust are valuable attributes. It is both appropriate and feasible for schools to develop children's strength and facility in these directions. But schools cannot do so if children lack day-to-day contact with children whose racial-ethnic identities differ from their own. The point here is not that contact per se can be counted on to produce interethnic acceptance. Recent studies show clearly that racial-ethnic boundaries function to organize patterns of social interaction in desegregated school settings (Singleton & Asher, 1979). Furthermore, racial-ethnic encapsulation is more prevalent among girls than boys (Rogers & Miller, 1981; Schofield & Francis, 1982) and hostility is

manifested more overtly on the playground than in classrooms (Rogers & Miller, 1981). The list of boundary conditions under which contact is likely to increase interethnic acceptance has grown increasingly longer (Cook, 1983; Stephan & Stephan, 1983).

On the other hand, and perhaps in response to the growing realization that they are needed, social scientists have begun to develop educational technologies that successfullly promote increased interethnic acceptance (Aronson et al. 1978; Cohen & Roper, 1972; Cook, 1982; DeVries, Edwards, & Slavin, 1978; Johnson, 1975; Rogers, Hennigan, & Miller, 1981; Sharan & Sharan, 1976; Slavin, 1978; Serow & Solomon, 1979). Though these procedures differ in their details, the common thread among them is their use of structured cooperative interaction in small groups, whether in conjunction with the curriculum or on the playground. Meta-analyses of their use not only show consistent and substantial benefit to interethnic acceptance, but improved academic mastery when coordinated with academic curriculum materials (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Johnson, Johnson, & Maruyama, 1983).

In summary, it is appropriate for schools to be concerned with childrens' development of effective and constructive interpersonal skills. The capacity for interethnic acceptance, respect, and trust is an important aspect of intrapersonal development and requires the existence of desegregated schools. Among the various goals that might be achieved by desegregated schooling, increased interethnic acceptance most directly addresses the central concern of Brown, namely, the stigmatization

of blacks. Thus, I would argue that even if on the average the effect of desegregated schooling on academic achievement was shown to be zero, desegregated schooling is required if the issue of interracial acceptance is to be addressed.

Conclusion

Taken together, the desegregation studies that meet the NIE minimal criteria show some moderate academic benefit to black children when they attend desegregated schools. Although one reviewer finds a larger margin of benefit among studies with stronger designs (Crain & Mehard, 1978) most reviewers find that the magnitude of effect is smaller in studies with better research designs (e.g., Krol, 1978; St. John, 1975). My calculation of the magnitude of these effects translates into the rather trivial increase of about twenty points, on the typical SAT college entrance test which has a mean of 500 and a standard deviation of 100. Most studies of desegregation assess the effects of only a year of desegregated schooling. The likelihood, however, that twelve years of desegregated schooling will translate into an average gain of over 200 points (two standard deviations) on an SAT type of test seems low. Our own longitudinal data from Riverside California certainly argue against such a view (Gerard & Miller, 1975). On the other hand, the high likelihood that the same level of performance is evaluated more favorably by the external world if a black student attends a desegregated as opposed to a segregated school, must be added to this picture. Given equal grade point averages or achievement test scores, the

black student from a desegregated school is likely to be viewed as more capable and promising than his or her peer from a segregated school.

My analyses of these and other data argue that the ratio of black and white students per se is probably not a direct causal factor in producing the small positive effect that is found. The fact that the magnitude of benefit is greater in studies conducted in the sixties than in those of the seventies supports this interpretation. The higher expectations and greater resources available in the earlier era should have generated increased morale and greater disruption of the status quo, thereby breaking the system effects that ordinarily depress the academic mastery of black children. Thus, I am arguing that whatever the academic effects found, they are due to teachers and schools and only attributable to changes in the percentages of black and white students to the extent that such changes concomitantly change teachers and schools.

Given the school effects that have been described in earlier sections, one could argue that such results essentially argue against the desegregation of schools. Implying as they do that lower class minority schools can be effective, education administrators should simply make the changes necessary to see that all such schools function effectively. Such a suggestion is not without merit, but is not easy to implement. When new teachers are brought into such schools to replace old ones, the normative structure exerts its influence on them, making them similar in outlook and practice to those they replaced. Such

systems of norms can continue to show their effects, even when all the persons in the system have one by one been replaced (Jacobs & Campbell, 1961). As new persons come into the system they too adopt the old norms and in turn, transmit them to still newer replacements.

For these reasons a change in the black child's school environment is more easily achieved by moving him or her to a more middle class school, than by attempting to change the school currently being attended. Middle class schools, being more likely to be high achieving schools, are less likely to have these debilitating systems of norms. Such a change can also give the minority student a sense of a fresh start.

In conclusion, the fact that school desegregation does not depress the academic performance of black children, but instead is moderately positive in its effect, (and as revealed in other reviews, does not adversely affect that of white children), means that if there are other compelling reasons to desegregate schools, consideration of academic achievement provides no deterrence. Because racially mixed schools are necessary if effective programs for increasing intergroup acceptance are to be applied, school desegregation should be encouraged.

References

Armor, D.J., The evidence on desegregation and black achievement. Paper, commissioned by the National Institute of Education. Washington D.C., 1983.

Aronson, E., Bridgeman, D., & Geffner, R. Interdependent interactions and prosocial behavior. A Journal of Research and Development in Education, 1978, 12, 16-27.

Banks, W.C. White Preference in Blacks: A Paradigm in Search of a Phenomenon." Psychological Bulletin, 1976, 83, 1179-1186.

Baughman, E.E. Black Americans: A Psychological Analysis. New York: Academic Press, 1971.

Beez, V.W., Influence of biased psychological reports on teacher behavior and pupil performance. In M. Miles & W.W. Charters (eds.) Learning in Social Settings. Boston: Allyn and Bacon, 1970.

Black, C. The lawfulness of the segregation decisions. Yale law Journal, 1960, 69 421-430.

Bowles, S. & Gintis, H. Schooling in Capitalist America. New York: Basic Books, 1976.

Brand, E.S., Ruiz, R.A., and Padilla, A.M. Ethnic Identification and Preference. Psychological Bulletin. 1974, 81, 860-890.

Brink, W. & Harris, L. Black and White, New York: Simon and Schuster, 1969.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker School Social Systems and Student Achievement: Schools can make a Difference. New York: Praeger, 1979.

Brophy, J.E. & Good, T.L. Teachers' Communication of differential

expectations for children's classroom performance. Journal of Educational Psychology, 1970, 61, 367-374.

Campbell, D.T. Stereotypes and Perception of Group Differences. American Psychologist, 1967, 22, 817-829.

Carrigan, P.M. School desegregation via compulsory pupil transfer: Early effects on elementary school children Ann Arbor, Michigan: Ann Arbor Public Schools, 1969.

Clark, N. Development of Consciousness of Self and the Emergence of Racial Identification in Negro Pre-school children. Journal of Social Psychology, 1939, 10, 591-599.

Cohen, E. & Roper, S. Modification of interracial interaction disability: An application of status characteristics theory. American Sociological Review, 1972, 37, 643-657.

Cook, S. Motives in a conceptual analysis of attitude-related behavior. In W. Arnold and D. Levine (Eds.), Nebraska Symposium on Motivation. Lincoln, Nebraska: University of Nebraska Press, 1969.

Cook, S.W. Social Science and School Desegregation: Did we Mislead the Supreme Court? Personality and Psychology Bulletin, 1979, 5, 420-437.

Cook, S.W. Cooperative interaction in multi-ethnic contexts. In N. Miller & M.B. Brewer (Eds.), Groups in Contact: Desegregation. New York: Academic Press (in press).

Crain, R.L., & Mahard, R.E. School racial composition and black college attendance and achievement test performance. Sociology of Education. 1978 51, 81-101.

Crain, R.L. & Weisman, C.S. Discrimination, Personality, and Achievement: A Survey of Northern Blacks. New York: Seminar Press, 1972.

Crawford v. Board of Education, L.A. No. 30485, 1976, 17 C3d 280-310

Curtis, B. The effect of segregation on the vocational aspirations of negro students. Dissertation Abstracts, 1968, 29, 772.

DeBord, L.W., Griffin, L.J., and Clark, M. Race and Sex Influences in the Schooling Processes of Rural and Small Town Youth. Sociology of Education, 1977, 50, 85-102.

DeVries, D.L., Edwards, K.J., & Slavin, R.E. Biracial learning teams and race relations in the classroom: Four field experiments on teams-games-tournament. Report #230, Center for Social Organization of Schools, Johns Hopkins University, 1977.

Dworkin, R. Social Sciences and constitutional rights - the consequences of uncertainty. In R.C. Rist & R.J. Anson (Eds.), Education, Social Science, and the Judicial Process. New York: Teachers College Press, 1977.

Edmonds, R.R. Search for effective schools: The identification and analysis of the schools that are instructionally effective for poor children. Unpublished manuscript, Harvard University, 1976.

Entwisle, D.R. & Hayduk, L.A. Early Schooling: Cognitive and Affective Outcomes. Baltimore: Johns Hopkins University Press, 1982.

Epps, E.G. Impact of School Desegregation on Aspirations, Self-Concepts and Other Aspects of Personality. Law and contemporary Problems, 1975, 39, 300-313.

Epps, E.G. The Impact of School Desegregation on the Self Evaluation and Achievement Orientation of Minority Children. Law and Contemporary Problems, 1979, 43, 57-76.

Falk, W.W. School desegregation and the educational attainment process: Some results from rural Texas schools. Sociology of Education, 1978, 51, 282-288

Fisher, J.E. An Exploration of the Effects of Desegregation on the Educational Plans of Negro and White boys. Dissertation Abstracts, 1971, 31, 5548

Fraser, R.W. Behavioral and attitudinal differences between teachers in desegregated classrooms. Unpublished doctoral dissertation, University of Southern California, 1981.

Gerard, H.B., & Miller, N. School Desegregation. New York: Plenum, 1975.

Glass, G.V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Glass, G.V. Integrating findings: The meta-analysis of research. Review of Research in Education, 1977, 5, 351-379.

Good, T.L. Which pupils do teachers call on? Elementary School Journal, 1970, 70, 190-198.

Gottfredson, D.C. Personality and persistence in education: A longitudinal study. Paper presented at the Annual Meeting of the American Psychological Association, Montreal, Canada,

Hall, J.A. & Wiant, H.V. Does school desegregation charge

occupational goals of negro males? Journal of Vocational Behavior, 1973, 3, 175-179

Hauser, R.M. On "a reconceptualization of school effects." Sociology of Education, 1978, 51, 68-72.

Heim, J.J. What research says about improving student performance. Albany: The University of the State of New York, the State Education Department, Bureau of School Programs Evaluation, March 1973.

Heyns, B. Summer Learning. New York: Academic Press, 1978.

Hoffman, D. & Cohen, E.G. An Exploratory study to determine the effects of generalized performance expectations upon activity and influence of students engaged in a group simulation game. Paper presented at the Annual meetings of the American Educational Research Association, Chicago, 1972.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. Meta-analysis: Cumulating Research Findings Across Studies. Beverly Hills: Sage, 1982.

Jacobs, R.C. & Campbell, D.T. The perpetuation of an arbitary tradition through several generations of a laboratory microculture. Journal of Abnormal and Social Psychology, 1961, 62, 649-658.

Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B., Michelson, S. Inequality. New York: Basic books, 1972.

Johnson, D.W., Cooperativeness and social perspective taking. Journal of Personality and Social Psychology, 1975, 31,

241-244.

Johnson, E., Gerard, H., & Miller, M. Teacher influences in the segregated classroom. In H.B. Gerard & N. Miller (Eds.), School Desegregation, New York: Plenum Press, 1975.

Johnson, D.W., Johnson, R.T., and Maruyama, G. Effects of cooperative learning: A meta-analysis. In N. Miller & M.B. Brewer (Eds.), Groups in Contact: Desegregation. New York: Academic Press (in press).

Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D. & Skon, L. Effects of cooperative competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 1981, 89, 47-62.

Katz, I. Factors influencing Negro performance in the desegrated school. In M. Deutsch, I. Katz & A.R. Jensen (Eds.), Social Class, Race, and Psychological Development. New York: Holt, Rinehart, and Winston, 1968.

Kenny, D.A. Cross Lagged Panel Correlations: A Test for Spuriousness. Psychological Bulletin, 1975, 82, 887-903.

Kluger, R. Simple Justice. New York: Vintage, 1977.

Knapp, N., and Hammer, E. "Racial Composition of Southern Schools and Adolescent Educational and Occupational Aspirations and Expectations." Paper presented at annual meeting of the Association of Southern Agricultural Workers, Memphis, January 1971.

Krol, R.A. "A Meta-Analysis of comparative Research on the Effect of Desegregation on Academic Achievement. "Unpublished doctoral dissertation, Western Michigan University, 1978.

Kulik, J.A., cohen, P.A., & Ebeling, B.J. Effectiveness of programmed instruction in higher education. Educational Evaluation and Policy Analysis, 1980, 2, 51-64.

Kulik, J.A., Kulik, C.-L.C. & Cohen, P.A. Effectiveness of computer based college teaching. Review of Educational Research 1980, 50, 525-544.

Kulik, J.A., Kulik, C.-L.C. & Cohen, P.A. Research an audio-tutorial instruction. Research in Higher Education, 1979a, 11, 321-341.

Kulik, C.-L.C. Schwalb, B.J., Kulik, J.A. Programmed instruction in secondary education. Journal of Educational Research, (in press).

Lippitt, R. & Gold, M. Classroom social structure as a mental health problem. Journal of Social Issues, 1959, 15, 40-49.

Maruyama, G. & Miller, N. Reexamination of normative influence processes in desegregated classrooms. American Educational Research Journal, 1979, 16, 273-284.

Maruyama, G. & Miller, N. The relation between popularity and achievement: A longitudinal test of the lateral transmission of value hypothesis. Unpublished paper, 1983.

Miller, N. Changing views about the effects of school desegregation: Brown then and now. In M.B. Brewer & B.E. Collins (Eds.), Scientific Inquiry and the Social Sciences San Francisco: Jossey-Bass, 1981.

Mussen, P.H. Differences Between the TAT Responses of Negro and White Boys. Journal of consulting Psychology, 1953, 17, 373-376.

Oakes, J. Classroom social relationships: Exploring the Bowles and Gintis hypothesis. Sociology of Education, 1982, 55, 197-212.

Pettigrew, T. Social evaluation theory: Convergences and applications. In D. Levine (Eds.), Nebraska Symposium on Motivation. (Vol. 15), Lincoln, Nebraska: University of Nebraska Press, 1967.

Porter, J.D.R., and Washington, R.E. Black Identity and Self-Esteem: A Review of Studies of Black Self-Concept, 1968-1978. Annual Review of Sociology, 1979, 5, 53-74.

Proshansky, H., and Newton, P. The Nature and Meaning of Negro Self-Identity." In M. Deutsch, I.Katz, and A.R. Jenson (Eds.) Social Class, Race, Psychological Development. New York: Holt, Rinehare and Winston, 1968.

Reniston, E.G. Levels of Aspiration of Black Students as a Function of Significant Others in Integrated and Segregated Schools. Dissertation Abstracts, 1973, 33, 7020-7021.

Rogers, M. & Miller, N. Quantitative and qualitative differences in peer selection among desegrated school children. Paper presented at the Annual Meeting of the American Psychological Association, Montreal, Canada, 1980.

Rogers, M. & Miller, N. The effect of School setting on cross racial interaction. Paper presented at the Annual Meeting of the American Psychological Association, Los Angeles, 1981.

Rogers, M., Miller, N., Henningan, K. Cooperative games as an intervention to promote cross-racial acceptance. American

Educational Research Journal, 1981, 18, 513-516.

Rosenholtz, S. & Rosenholtz, S.H. Classroom organization and the perception of ability. Sociology of Education, 1981, 54, 132-140.

Rubin, R.A., Maruyama, G., & Kingsbury, G.G. Self-esteem and educational achievement: A causal model analysis. Paper presented at the Annual Meeting of the American Psychological Association, New York, 1979.

Scheirer, M.A. & Kraut, R.E. Increasing educational achievement via self-concept charge. Review of Educational Research, 1979, 49, 131-150.

Schofield, J.W. Complementary and conflicting identities: Images and interaction in an interracial school. In S. Asher & J. Gottman (Eds.), The Development of Children's Friendship. New York: Cambridge University Press, 1980.

Schofield, J.W. & Francis, W.D. An observational study of peer interaction in racially mixed "accelerated" classrooms. Journal of Educational Psychology, 82, 74, 722-732.

Serow, R.C. & Solomon, D. Classroom climate and students intergroup behavior. Journal of Educational Psychology, 1979, 71, 669-676.

Sewell, W.H., Haller, A.O., & Portes, A. The educational and early occupational attainment process. American Sociological Review, 1969, 34, 82-92.

Sharan, S. and Sharan, Y. Small-group Teaching. Englewood Cliffs: Educational Technology Publications, 1976.

Simpson, C. Classroom structure and the organization of ability. Sociology of Education, 1981, 54, 120-132.

Singleton, L.C. & Asher, S.R. Racial integration and children's peer preferences: An investigation of developmental and cohort differences. Child Development, 1979, 50, 936-941.

Slavin, R.E. Student teams ad comparison among equals: Effects on academic performance and student attitudes. Journal of Educational Psychology, 1978, 70, 532-538.

St. John, N.H. The Effect of Segregation on the Aspirations of Negro Youth. Harvard Educational Review, 1966, 36, 284-294.

St. John N.H. School Desegregation outcomes for children New York: Wiley, 1975.

Stephan, W.G. School Desegregation: An Evaluation of Predictions Made in Brown vs. Board of Education. Psychological Bulletin, 1978, 85, 217-238.

Stephan, W.G. Blacks and Brown: The effects of school desegregation on black students. Paper commissioned by the National Institute of Education. Washington, D.C., 1983.

Stephan, W.G. & Stephan, C.W. The role of ignorance in intergroup relations: Increasing knowledge of ethnic outgroups in multi-ethnic classrooms. In N. Miller & M.B. Brewer (Eds.), Groups in Contact: Desegregation. New York: Academic Press (in press).

Summers, A.A. & Wolfe, B.L. Do schools make a difference? American Economic Review, 1977, 65, 639-52.

Tammivaara, J.S. The effects of task structure on beliefs about competence and participation in small groups. Sociology of

Education, 1982, 55, 212-222.

Thompson, E.W., & Smidchens, U. Longitudinal effects of school racial/ethnic composition upon student achievement. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Walberg, H.J. What makes Schooling effective? Contemporary Education Review, 1982, 1, 1-34.

Walberg, H.J. Desegregation and educational productivity. Paper, commissioned by the National Institute of Education, Washington D.C., 1983.

Weinberg, M. The relationship between school desegregation and academic achievement: A review of the research. Law and Contemporary Problems, 1975, 39, 240-270.

White, K. & Knight, J.H. School desegregation, SES, SEX, and the aspirations of Southern Negro adolescents. Journal of Negro Education, 1973, 42, 71-78.

Wilson, A.B. Residential segregation of social classes and aspirations of high school boys. American Sociological Review. 1959, 24, 836-845.

Wilson, W. Rank order of discrimination and it's relevance to civil rights priorities. Journal of Personality and Social Psychology, 1970, 15, 188-224.

Wortman, P.M. School desegregation and black achievement: A meta-analysis. Paper, commissioned by the National Institute of Education, Washington D.C., 1983.

Wylie, R.C. The Self-Concept: Theory and Research on Selected Topics (Vol. 2), Lincoln: University of Nebraska Press,

1979.

Yudof, M.G. Nondiscrimination and beyond: The search for principle in supreme court desegregation decisions. In W.G. Stephan & J.R. Feagin (Eds.), School Desegregation: Past, Present, and Future. New York: Plenum, 1980.

Appendix A

1) Type of Study

   a) non-empirical
   b) summary report

2) Location

   a) outside USA
   b) geographically non specific

3) Comparisons

   a) not a study of achievement of desegregated blacks
   b) multi-ethnic combined
   c) comparisons across ethnics only
   d) heterogeneous proportions minority in desegregated condition
   e) no control data
   f) no pre-desegregation data
   g) control measures not contemporaneous
   h) majority black in a segregated condition (unless the reviewer provides specific justification)
   i) varied exposure to desegregation (unless the reviewer provides a specific justification demonstrating that the variation in exposure time is not meaningful)

4) Study Desegregation

   a) cross-sectional survey
   b) sampling procedure unknown
   c) separate non-comparable samples at each observation

5) Measures

   a) unreliable and/or unstandardized instruments
   b) test content and/or instrument unknown
   c) dates of administration unknown
   d) different tests used in pre-tests and post-tests
   e) test of IQ or verbal ability

6) Data Analysis

   a) no pre-test means
   b) no post-test means, unless the author reported pre-test scores and gains
   c) no data presented
   d) The following will be rejected dependent upon the amount of information available for the reviewer to estimate values
      1. no pre-test standard deviations
      2. no post-test standard deviations
      3. no significance tests
      4. N's not discernable

It was decided that "excessive attrition" and "groups that are initially non-comparable" would not be used as criterion for rejection. In each case it was argued that the point at which the problem became an issue was extremely vague. It was felt that the project is better served by including studies exhibiting attrition and comparability problems and allowing individual reviewers to articulate these limitations. Using this criteria, 18 studies were selected which were deemed acceptable for inclusion in the project. These are:

Anderson, Lewis V. The effect of desegregation on the achievement and personality of Negro children. Unpublished doctoral dissertation, George Peabody College for Teachers, 1966. (University Microfilm 66-11, 237.

Baker, Jerome. A study of integration in racially imbalanced urban public schools. Syracuse, New York: Syracuse University Youth Development Center, Final Report, May 1977.

Bowman, Orrin H. Scholastic development of disadvantaged Negro pupils: A study of pupils in selected segregated and desegregated elementary classrooms. Unpublished doctoral dissertation, University of New York at Buffalo, 1973.

Carrigan, Patricia M. School desegregation via compulsory pupil transfer: Early effects on elementary school children. Ann Arbor, Michigan: Ann Arbor Public Schools, 1969.

Clark, El Nadel. Analysis of the difference between pre- and post-test scores (change scores) on measures of self-concept, academic aptitude, and reading achievement earned by sixth grade students attending segregated and desegregated schools. Unpublished doctoral dissertation, Duke University, 1971.

Evans, Charles L. Short term desegregation effects: The academic achievement of bused students 1971-1972. Fort Worth, Texas: Fort Worth Independent School District, 1973. (ERIC No. ED 086 759)

Iwanicki, E.F., & Gable R.K. A quasi-experimental evaluation of the effects of a voluntary urban/suburban busing program on student achievement. Paper presented at the Annual meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Klein, Robert Stanley. A comparative study of the academic achievement of Negro tenth grade high school students attending segregated and recently integrated schools in a metropolitan area in the south. Unpublished doctoral dissertation. University of South Carolina, 1967.

Laird, M.A. & Weeks, G. The effect of busing on achievement in reading and arithmetic in three Philadelphia schools. Philadelphia, Pennsylvania: The School District of Philadelphia, Division of Research, 1966.

Rentsch, George J. Open-enrollment: An appraisal. Unpublished doctoral dissertation, State University of New York, Buffalo, 1967.

Savage, L.W. Arithmetic achievement of black students transfering from a segregated junior high school to an integrated junior high school. Unpublished masters thesis, Virginia State College, 1971.

Sheehan, Daniel S. "Black Achievement in a desegregated school district." Journal of Social Psychology, 1979, 107, 165-182

Slone, Irene W. The effects of one school pairing on pupil achievement, anxieties and attitudes. Unpublished doctoral dissertation, New York University, 1968.

Syracuse City School District. Study of the effects of integration -- Washington Irving and Host Pupils. Hearing held in Rochester, New York, September 16-17, U.S. Commission on Civil Rights,

Thompson, E.W., & Smidchens, U. Longitudinal effects of school racial/ethnic composition upon student achievement. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April, 1979).

Van Every, D.W. Effects of desegregation on public school groups of sixth graders in terms of achievement levels and attitudes toward school. Doctoral dissertation, Wayne State University, 1969. Dissertation Abstracts International, 1969. (University Microfilms No. 70-19074)

Walberg, Herbert J. An evaluation of an urban-suburban school bussing program: Student achievement and perception of class learning environments. Paper presented at the annual meeting of the American Educational Research Association, New York, New York, February 1971.

Zdep, Stanley M. "Educating disadvantaged urban children in suburban schools: An evaluation." Journal of Applied Social Psychology, 1971. (ERIC No. ED 053 186 TM 00716)